

DOOMSDAY, BISHOP USSHER AND SIMULATED WORLDS.*

Alasdair M. Richmond

ABSTRACT

This paper attempts three tasks in relation to Carter and Leslie's Doomsday Argument. First, it criticises Timothy Chambers' 'Ussherian Corollary', a striking but unsuccessful objection to standard Doomsday arguments. Second, it reformulates the Ussherian Corollary as an objection to Bradley Monton's variant Doomsday and Nick Bostrom's Simulation Argument. Finally, it tries to diagnose the epistemic/metaphysical problems facing Doomsday-related arguments.

D) Carter and Leslie's Doomsday Argument.

Human extinction in the near future seems a sadly possible eventuality. While all-out nuclear war seems less likely than it once did, other threats to our survival remain plentiful. Under such conditions of uncertainty, how should we assess our species' prospects? One intriguing answer to this question is known as the Doomsday Argument (DA). Unlike prophetic risk-assessments based on straightforward projections of existing (or foreseeable) threats, DA instead looks at what probabilistic reasoning might suggest about our likely location in human history. In particular, rather than try to predict the course of the future, DA aims to determine what sort of future would make our current location probable. DA concerns, not necessarily what will happen, but the sort of future that we ought to believe in.

Using only some plausible assumptions about probabilistic sampling and our likely number of ancestors, DA urges us to increase our personal probability for imminent human extinction. DA was first aired in a 1983 public lecture by Brandon Carter¹ and has subsequently been much championed by John Leslie.² Carter-Leslie DA applies

*Very many thanks to John Leslie for helpful comments on an earlier draft of this article.

¹A version of Carter's lecture was published as 'The Anthropic Principle and its Implications for Biological Evolution', *Philosophical Transactions of the Royal Society of London*, Series A, Vol. 310, 1983, pp. 347-63. However, the published version alludes to DA only tangentially.

²The essential DA work is Leslie's *The End of the World: The Science and Ethics of Human Extinction*, (London, Routledge, 1996, expanded edition 1998). Particularly salient for DA-critics are Leslie's (1998, pp. 206-36) replies to objections. Also recommended is Nick Bostrom's *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, (London, Routledge, 2002), esp. pp.

Bayesian reasoning to hypotheses about the total human population, conditional on our approximate birth-rank in history. Estimates of past population suggest that contemporary humans have birth-ranks around sixty billion. Considering ourselves as random humans, our living now is fairly unremarkable if extinction looms but we are very unusual humans if humanity keeps growing. If the all-time human total proves to be (e.g.) six trillion, then less than 1% of humans will have had birth-ranks below sixty billion. However, if Doom occurs tomorrow, almost all human birth-ranks will be below ours. If humanity survives at or near its present level of population, our birth-ranks will quickly become unusually low. Our current location appears quite likely under the hypothesis of imminent extinction but looks rather less likely under hypotheses of long-term survival. If we should favour those hypotheses that make their explananda more likely rather than less, we should accordingly be inclined to favour the hypothesis of impending Doom.

Carter-Leslie DA is a subjectivist Bayesian argument, whose explanandum is our having the birth-ranks we do and whose explanans is a theory about future human population. DA tries to derive degrees of credence for our location, based on assumptions about the prior probabilities of certain population hypotheses and the likelihood these hypotheses bestow on our living hereabouts in history. DA probabilities reflect personal degrees of belief and not (e.g.) the limiting frequency with which species like ours die out. Before it can generate its conclusions, DA needs some assumptions about how to bestow prior probabilities on population-hypotheses and likelihoods on birth-ranks. Carter-Leslie DA usually uses a principle of indifference to generate prior probabilities, e.g. if you assume a maximal human population of x , give all population-hypotheses a prior of $1/x$. Likewise, Carter-Leslie DA normally gives each birth-rank a likelihood which is inversely proportional to the total population postulated, e.g. if you think the all-time human population will be 100 billion, your birth-rank should have a likelihood of $1/100$ billion.

After Leslie (pp. 199-200), we can illustrate DA thus: your name is written on a slip of paper and placed in an urn. To keep things simple, let's only use two hypotheses about the urn, i.e. $H1$ = "the urn holds ten names" and $H2$ = "the urn holds

89-126. My 'Recent Work: The Doomsday Argument' (*Philosophical Books*, Vol. 47, 2006, pp. 129-42) offers a survey of DA, its forerunners and relatives.

a thousand names”. You give H1 and H2 prior probabilities of 0.05 and 0.95 respectively. Names are drawn from the urn, without replacement. Your name is drawn third. Should your name appearing so early affect your probabilities? Using Bayes’ Theorem, plus the above prior and likelihood assumptions, transforms our prior of 0.05 into a posterior probability around 0.84.³ If your name appears early, the urn probably holds few names. Likewise, your birth-rank confers higher conditional probability on the all-time human population being relatively small. If we should prefer explanations which confer greater conditional probability on our birth-ranks, we should assume that time will probably not add many more people to the human total, i.e. we should believe that humanity probably won’t survive much longer.

DA variants are now legion. In an unusual counter to Carter-Leslie DA, Bradley Monton tries to reconstruct DA without birth-rank data.⁴ Nick Bostrom’s DA-related Simulation Argument says: if we are functionalists who think advanced civilisations will command much greater computing resources than ours, we should believe we’re probably simulated minds.⁵ Against such DA exotica, this paper deploys Timothy Chambers’ anti-DA ‘Ussherian Corollary’.⁶ Chambers’ argument may fail against traditional DA but it’s an instructive failure which reveals much about DA’s evidential basis and can still pose problems for Monton and Bostrom.

II) A Young Earth.

Timothy Chambers’ ‘Ussherian Corollary’ (UC) is a striking parody of Doomsday reasoning, named after Bishop Ussher’s 4,004 BC date for the Creation. Chambers (p. 446) says DA “entails a parallel Ussherian moral: that we have systematically underestimated the chance that the human race began fairly recently”. Thus, DA favours recent Creation as much as Doom Soon, i.e. we are more likely to find ourselves living now if humanity is new rather than old. As Chambers thinks the Urn

³ $P(H1 | e) = \frac{P(e | H1) P(H1)}{P(e | H1)P(H1) + P(e | H2) P(H2)} = \frac{0.1 \times 0.05}{(0.1 \times 0.05) + (0.001 \times 0.95)} \approx 0.84$

⁴ ‘The Doomsday Argument Without Knowledge of Birth Rank’, *The Philosophical Quarterly*, Vol. 53, 2003, pp. 79-82.

⁵ ‘Are We Living in a Computer Simulation?’, *The Philosophical Quarterly*, Vol. 53, 2003, pp. 243-255.

⁶ ‘Do Doomsday’s Proponents Think We Were Born Yesterday?’, *Philosophy*, Vol. 76, 2001, pp. 443-50.

story is formally compatible with UC and DA, he maintains that anyone who finds UC unacceptable should also reject DA. However, even if we grant Chambers this formal parity, it doesn't follow that he has refuted DA. Five replies to UC follow. Replies 1-4 are peripheral but #5 suggests a real disanalogy between DA and UC.

Reply 1: So what if DA and UC are symmetrical? We can all seem improbable relative to *some* set of descriptors, and probabilistic arguments often fail for someone in their reference class. (Somebody has to be the winner of even a billion-ticket fair lottery.) Maybe we can accept DA, but also conclude that we are simply unusual humans relative to UC's descriptors. However, if you refuse to be moved by apparent unlikelihoods of location, DA won't appeal to you anyway.

Reply 2: UC suggests the more recently history began, the likelier our location becomes. Thus, Chambers' parody doesn't support an *Ussherian* Corollary but a more radical Russellian one, e.g. the Earth began five minutes ago with six billion people already *in situ*. However, this reply suggests UC merely errs in putting Creation's date too long ago. A shorter past makes our present location likelier still.

Reply 3: Far from refuting DA, UC makes Doom *more* likely. If Ussher was right, Adam and Eve had at least six billion descendents in c. 6,000 years, implying steep and sustained population-growth. DA suggests that the steeper the population curve, the quicker our birth-ranks become unusual. (If this were a nanosecond-old 'Russellian' world, the rapid appearance of six billion people would mean a near-vertical population-curve. Our birth-ranks would be *very* early if this trend continued even for microseconds.) So, UC gives us more reason to accept DA. However this reply merely reinforces Chambers' alleged parity between UC and DA.

Reply 4: DA may seem to favour literal *doom*, (i.e. extinction) but it could equally suggest that our descendents evolve into something radically different from ourselves. Likewise, UC may not favour recent *Creation*, as opposed to a recent cognitive breakthrough that puts contemporary humans in a different reference class from past ones.⁷ However, we could equally 'explain' our present location on the hypothesis that all our predecessors were mindless hulks who went through life untroubled by consciousness. Such a 'Zombie' Corollary scarcely seems to improve on UC.

⁷Cf. Julian Jaynes' *The Origins of Consciousness in the Breakdown of the Bicameral Mind*, (Princeton, Princeton University Press, 1976).

Reply 5: Replies 1-4 don't address the alleged parity between UC and DA. Even if UC and DA are formally symmetrical, their respective evidential bases are not. UC might threaten DA *if* DA ignored all population data and derived our likely future purely from the fact that we exist now. However, birth-rank data underpin DA as much as Bayes' Theorem does. In emphasising DA's formal properties and downplaying its evidential basis, Chambers risks depicting DA as an *a priori* argument or one for which evidential considerations are secondary. DA offers a scaling-factor for updating our probabilities. Any such argument can act only as a lever (or force-multiplier); it cannot dictate probabilities or credences on its own. We may think UC and DA formally symmetrical but find Doom Soon much more likely than recent Creation. If our probability for recent Creation is sufficiently low before invoking UC, it may still be low afterwards.⁸

We could remove the evidential asymmetry between UC and DA, e.g. by discounting all evidence of our population-history. Observers who were equally uncertain about past and future might find DA and UC symmetrical evidentially as well as formally. However, even if DA and UC look similar from an evidential standpoint unlike ours, this need not shed any light on DA's validity for us. Our knowledge of the past is conditional only on such comparatively uncontroversial assumptions as that archaeology and radiocarbon-dating uncover real traces of the past and not experimental artefacts. Supporters of DA can equally "believe that empirical methods (e.g. radiometric dating of fossils) are completely apt for discerning the age of the human race", (Chambers, p. 447).

Chambers imagines some future theory undercutting DA and so restoring evidential parity between DA and UC, e.g. that "sufficient cosmological data will ultimately rule out 'Doom Soon' possibilities", (Chambers, p. 447). Presumably, he has in mind something like the 'Final Anthropic Principle', i.e. "Intelligent information-processing must come into existence in the Universe, and, once it comes into existence, it will never die out".⁹ It isn't clear how far present applications of

⁸Chambers (p. 446, fn. 10) alludes to this problem but seemingly misses its import for his own argument.

⁹John D. Barrow and Frank J. Tipler, *The Anthropic Cosmological Principle*, (Oxford, Oxford University Press, 1986), p. 23.

DA are threatened by the prospect of such hypothetical future theories. To undercut DA, the necessity of *human* survival would have to be established but even the Final Anthropic Principle only claims that life's survival is necessary. Having baptised anthropic reasoning, Carter cautioned against its anthropocentric use: "The same self-selection principle would be applicable by any extra-terrestrial civilisation that may exist", (Carter, p. 348).¹⁰ Evidential symmetry between DA and UC needs more than hints that life's survival may be necessary; it needs census-returns from the future.

III) A Place in the Sun.

DA generalises from our existing evidential basis and does not incite us to discard any portion thereof. DA and UC are not just asymmetrical in terms of evidence but, more importantly, asymmetrical in terms of how they *treat* evidence. DA does *not* urge us to choose whichever reference class makes our location most probable, regardless of any violence thereby done to our existing evidential basis. If this were DA's method, DA would scarcely need refuting. Any such evidence-discounting DA is apparently a straw man of Chambers' invention.

However, an Ussherian might say: DA birth-ranks are inferred from defeasible empirical data, whereas Bishop Ussher's population-estimates derive from a revealed text. Furthermore, Ussher's hypothesis can be made compatible with experience.¹¹ However, anyone who accepts a different revelation could use the same reasoning to support any number of evidence-discounting inferences. Why stop with UC and a Young Earth? Discounting empirical evidence purely so that our location may appear probable can have highly counter-intuitive consequences. Likewise with contriving compatibility between our evidential basis and our test-hypothesis.

Consider a 'Solarian Corollary' (SC). In 1714, Tobias Swinden argued that Hell is located in the Sun.¹² Besides giving Hell its requisite heat and size, his hypothesis

¹⁰Bostrom (2002, p. 50) dismisses the Final Anthropic Principle as "antithetic to Carter's anthropic principle" and "pure speculation".

¹¹Philip Gosse's *Omphalos: An Attempt to Untie the Geological Knot* (London, John Van Voorst, 1857) argued that any created Earth that contained true archetypes for subsequent life would have to bear false traces of a non-existent past, e.g. growth-rings in Eden's trees, a navel for Adam, etc. Gosse thought the unfalsifiability of his hypothesis was an inevitable hallmark of any creationist hypothesis.

¹²*An Enquiry into the Nature and Place of Hell*, (London, Thomas Astley, 2nd edition, 1727).

had other advantages, e.g. restoring Hell to the nadir of (a Copernican) creation. Swinden judged that only a solar Hell could hold the hundred thousand million (or more) souls who would comprise the total of the damned. Swinden's (10^{11}) damned out-weigh the (6×10^9) now living. Let's assume the saved are negligible and quantify over observers *sub specie æternitatis*. Life is short but damnation eternal. If we should favour our residing in the largest and/or most enduring soul-population, as random souls we're more likely damned in the Sun than alive on Earth. If Hell ever holds (10^{11}) souls, six billion souls now live on Earth, and we aren't sure where we are, our damnation is around 94% probable.¹³ (In such cases, we can derive our probabilities directly from the relative sizes of populations. Section 5 considers a probabilistically similar case.)

SC can even be made compatible with experience. Maybe: 1) damnation is tolerable until grasped as such, 2) the damned only grasp their situation gradually, or 3) others' *apparent* toleration of Hell is illusory. However, SC is self-frustrating. If this is Hell, nor am I out of it, many of my empirical beliefs are wrong, (e.g. my belief that people die out of this world). SC may make our location appear probable under one set of descriptors but disregards the associated epistemic violence done to our existing empirical beliefs. SC requires that we have somehow correctly grasped relative population-sizes while being deluded about our location. Any preference Solarians show for our population-beliefs is arbitrary. Likewise, UC requires that generations of scribes have preserved a true history of the world but that all empirical evidence of an ancient Earth is systematically false. Such pervasive deception would imperil almost all empirical data. So, another UC moral: respect the epistemic dimension of test-hypotheses and place more stringent requirements on them than simply demanding that they can be made consistent with experience.

IV) Doomsday Without Birth-Ranks?

Even if Chambers fails to despatch Carter-Leslie DA, UC may threaten some DA offshoots. DA discussions often debate the merits of two different rules for relating probabilities and populations. Some DA supporters invoke a Self-Sampling Assumption (SSA): "Observers should reason as if they were a random sample from the set of all observers in their reference class," (Monton, p. 80). Many DA-critics

¹³Prob. (Damned) = $\frac{\text{Total damned}}{\text{Total (damned + living)}} = \frac{10^{11}}{10^{11} + (6 \times 10^9)} = (100/106) \approx 94.34\%$

argue for ‘Doom Later’ shifts using a counterbalancing ‘Self-Indication Assumption’ (SIA): “Finding you exist gives you reason to think that there are many observers”, (Monton, p. 81).¹⁴ Even if SSA alone yields a ‘Doom Soon’ probability-shift, applying SIA *and* SSA yields no overall shift in probabilities conditional on birth-ranks. SIA seems attractive, not least because it reinforces the intuition that having a certain birth-rank should have no bearing on humanity’s longevity.

However, just because SIA undercuts DA it doesn’t follow that SIA is unproblematic. SIA makes no reference to birth-ranks and Bostrom argues that using SIA *alone* has absurd consequences. His ‘Presumptuous Philosopher’ story (2002, pp. 124 ff.) imagines cosmologists who have two (initially equiprobable) candidate ‘Theories of Everything’. Theories A and B postulate totals of 200 billion and 200 trillion observers respectively. Just before an experiment to decide between A and B, up pops a Presumptuous Philosopher, who says: your experiment is unnecessary – SIA suggests your existence is a thousand times more likely on B so your credences should favour B accordingly. Such reasoning seems counter-intuitive.

Monton (2003) defends SIA by arguing that DA doesn’t need birth-rank data. He reconstructs DA thus: H1 and H2 postulate total populations of 200 billion and 200 trillion humans respectively, and receive respective priors of 0.05 and 0.95. Let’s assume we also know proposition R, namely “I have birth rank 60 billion”. SSA implies $P(R | H1) = 1/200$ billion and $P(R | H2) = 1/200$ trillion. By Bayes’ Theorem, learning R transforms $P(H1) = 0.05$ into $P(H1 | R) = 0.98$.¹⁵ However, SIA suggests H2 makes our existing a thousand times more likely than H1 and the counterweight of this increased likelihood restores our original priors.

Keeping the above hypotheses and priors, how might DA run without birth-ranks? Monton’s DA uses a birth-rank surrogate property k , e.g. “being alone in 323 Main Street in Lexington, Kentucky, from 20:41 to 20:42 GMT on April 9, 2002”, (Monton, p. 80). Let’s assume H1 and H2 agree on the number of humans who possess k . Let ‘ K ’ be the proposition that someone has k . $P(K)$ is independent of

¹⁴A version of SIA seems to have been first proposed in Dennis Dieks’s ‘Doomsday - or the Dangers of Statistics’, *The Philosophical Quarterly*, Vol. 42, 1992, pp. 78–85.

¹⁵
$$P(H1 | R) = \frac{P(R | H1) P(H1)}{P(R | H1)P(H1) + P(R | H2)P(H2)} = \frac{(1/200 \text{ billion} \times 0.05)}{(1/200 \text{ billion} \times 0.05) + (1/200 \text{ trillion} \times 0.95)} \approx 0.98$$

whether H1 or H2 obtains, i.e. $P(K)$ is not conditional on overall population-size. (Monton needs this assumption, otherwise your learning K would support either H1 or H2.) At 20:42 GMT on April 9, 2002, you learn that someone (i.e. you) possesses k and hence that K is true. Now suppose 'M' is the proposition that *you* have property k , and you know M. Using SSA, $P(M | H1)$ is a thousand times greater than $P(M | H2)$, i.e. $P(M | H1) = 1/200$ billion and $P(M | H2) = 1/200$ trillion. Applying Bayes' Theorem to the above priors and likelihoods, $P(H1 | M) = 0.98$. Monton concludes that SSA can generate DA without birth-ranks, hence SSA and SIA have equally implausible consequences.

However, Darren Bradley thinks Monton's DA implicitly requires birth-ranks.¹⁶ Monton needs restrictions on admissible properties k because some properties will covertly encode birth-rank data. If we learn "there existed a 500th President of the United States", (Bradley, p. 96), this will favour H2's larger population. Monton's problem is that the choice of properties must be restricted so that k must be instantiated in the smaller population "otherwise K would shift the probabilities of the hypotheses", (Bradley, p. 97). So, when you learn K , you also know you belong in H1's smaller population and so know your approximate birth-rank.

Invoking UC supports Bradley's conclusion that DA needs birth-rank information and hence that SIA is not demonstrably more plausible than SSA. Even if Monton deflects Bradley's criticism by finding a rank-independent property k , we can make an Ussherian temporal mirror-image of Monton's DA: M has higher likelihood if the *past* population is small too. Thus, $P(M | H1) > P(M | H2)$ if H1 and H2 postulate the same number of present humans, but a small and large past humanity respectively. Monton's DA risks being formally symmetrical between past and future, with the twist that Monton cannot invoke birth-rank data. Using empirical evidence to distinguish between H1 and H2 would disclose our approximate birth-ranks. (Worse, given a property k which is equally likely to be instantiated among the living or the damned, SIA and Monton's argument might support a Solarian Corollary as well.) The chief objection to UC is that DA needs birth-rank data. Without such evidence, DA is no more compelling than other highly counter-intuitive attempts to make our

¹⁶'No Doomsday Argument Without Knowledge of Birth Rank: A Defense of Bostrom', *Synthese*, Vol. 144, 2005, pp. 91-100.

apparent location appear probable. So, a dilemma: Monton's DA either invokes birth-rank data or it succumbs to UC. Monton (p. 80) does specify that H1 and H2 "should be understood as agreeing on the number of humans that exist up to now and into the short term future". However, this is a stipulation, not an evidence-driven requirement.

V) The Simulation Argument.

Bostrom proposes a novel variation on DA. He argues thus: if functionalism is correct and being conscious *is* simply running a correctly-configured programme, advances in computing might one day allow computer-simulated minds to out-number carbon-based minds. (After Brian Weatherson, let's call fully-conscious simulations 'Sims'.¹⁷) Sufficiently fine-grained Sims might have experiences indistinguishable from those of non-Sims. Technological forecasts suggest we are now nowhere near the theoretical maxima for computing power or efficiency. Advanced civilisations might run many Sims. If we believe Sims will out-number embodied minds, and we aren't sure which we are, we ought to bet we're Sims:

If we knew that a fraction x of all observers with human-type experiences live in simulations, and we have no information to indicate that our own particular experiences are any more or less likely than other human-type experiences to have been implemented *in vivo* rather than *in machina* then our credence that we are in a simulation should equal x . Bostrom (2003, p. 249)

Thus, if we think Sims will ever out-number non-Sims, we should believe we're probably screen-savers on some posthuman equivalent of a desk-top.

Bostrom is *not* arguing that we're probably Sims. Rather, his conclusion is a disjunction. If we accept a functionalist conception of the mind and think future computing-power may greatly outstrip ours, we should assume either: a) few posthumans exist, b) posthumans run few Sims or c) we're probably Sims. (Cf. Reply 4 to UC.) If we create Sims ourselves, this would tell against a) or b) and in favour of c), i.e. if we ever run Sims, this would suggest we're probably Sims ourselves and we live in a nested hierarchy of simulated simulators: "If we do go on to create our own ancestor-simulations, ... we would therefore have to conclude that we live in a simulation. Moreover, we would have to suspect that the posthumans running our simulation are themselves simulated beings; and their creators in turn may also be simulated beings," (Bostrom, 2003, p. 253).

¹⁷ 'Are You a Sim?', *Philosophical Quarterly*, Vol. 53, 2003, pp. 425-31.

However, while Bostrom's Simulation Argument is related to DA, Bostrom rejects DA. In particular, he rejects the idea that we can treat ourselves as randomly-drawn humans and use indifference principles to create priors for population-hypotheses. He maintains that we can't consider ourselves randomly-selected humans because we have some relevant information about our place in history, i.e. we know roughly how many people preceded us.¹⁸ (Again, DA can't easily dispense with birth-rank data, on pain of falling victim to Chambers' UC.) Instead of the indifference principle behind Carter-Leslie DA, Bostrom (2003, p. 250) uses an apparently weaker 'bland indifference principle', which only counsels indifference "between hypotheses about which observer one is, when one has no information about which of these observers one is". Adapting Bostrom's example (2003, p. 250): suppose you know that $x\%$ of people have a certain strand of junk DNA. You know of no experiential difference between having and not having this strand. Short of getting a full gene-assay, you should give the hypothesis that you have this strand $x\%$ probability. (As with the Solarian Corollary, the Simulation case can derive its probabilities of location directly from the relative population-sizes involved.) Likewise, if we think that $x\%$ of observers are Sims and we have no evidence bearing on whether or not we are Sims, we should give our Sim-hood $x\%$ probability.

Bostrom's Simulation Argument assumes that the most salient reference class is composed of those observers that, for all we can tell, we might be.¹⁹ However, there are more ways than one of considering our location and some of them lead to conclusions at variance with Bostrom's. If we are Sims, we must occupy a location somewhere in a simulation-hierarchy. What sort of location? Well, we have some

¹⁸Two brief observations on indifference and DA: 1) Bostrom seemingly thinks birth-rank data invalidate DA's indifference principle. Why can't Carter-Leslie DA simply apply to birth-rank data the standard Bayesian strategies for dealing with old evidence? See Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, (La Salle, Open Court, 2nd edition, 1993), pp. 403 ff. (Bostrom (2002, p. 97) is sceptical of the *general* applicability of such 'counterfactual' solutions to old evidence problems.) 2) Even if Bostrom's bland indifference principle is superior to standard indifference principles, DA may not need such principles. Darren Bradley and Branden Fitelson ('Monty Hall, Doomsday and Confirmation', *Analysis*, Vol. 63, 2003, pp. 23-31) propose confirmation-theoretic DA that greatly weaken, or entirely bypass, indifference assumptions.

¹⁹Strictly, Bostrom thinks this reference class is an idealisation. His (2002, pp. 159-183) argues that the *truly* salient reference class partitions over observer-moments.

evidence bearing on our hierarchy-position. If we are Sims, we are apparently not simulators. At present, we lack both the hardware and the programming know-how for creating suitably fine-grained approximations of human consciousness. If we take Bostrom's simulation-hierarchy seriously, we have to conclude that ours is the one and only level in the hierarchy that lacks descendents. If we have no reason to think such unsimulating locations are probable, we should look with suspicion on any hypothesis that makes our location seem unusual. (Bostrom cannot assume that only non-simulating levels can support life, otherwise whence our simulators?)

Chambers claims the 'Urn Model' that yields high probabilities for our having few descendents can also generate a high probability for our having few ancestors. UC fails because it treats two evidentially very different situations (i.e. the past and the future) identically. However, a UC variant might run in the Simulation case, because any numbers for ancestral simulation-levels must needs be speculative. A successorless-level is only a probable location if being successorless is the norm, but that can only be so if no simulation occurs, i.e. only one level exists, bereft of ancestors and successors. This reasoning *could* be supported with a DA-style indifference-principle, e.g. if L is our maximal total of simulation-levels, our prior probability of occupying a given level equals $1/L$. However, all our anti-hierarchy conclusion needs is the weaker assumption that the *likelihood* of our occupying a given level declines as the total of levels increases. Thus, if you have two hypotheses A and B about your hierarchy, where A postulates more levels than B, you ought to prefer B, i.e. if your level is number x , $P(L_x | B) > P(L_x | A)$. Because this argument issues in a likelihood-ratio, it requires neither priors, exact numerical likelihoods nor indifference principles. That the likelihood of our occupying a given level declines as the total of levels rises is not an *a priori* claim but an empirically-driven assignment of likelihoods based on Bostrom's assumption that simulation costs increase with the total of levels simulated: "One consideration that counts against the multi-level hypothesis is that the computational cost for the basement-level simulators might be prohibitively great", (2003, p. 253).

Suppose we consider a more complicated simulation-hierarchy than a straightforward linear chain of simulation-levels. What if hierarchies can branch? Then we potentially have to consider our location *qua* observer, Sim-level or Sim-world. In a branching hierarchy, each level may run many simulated worlds, each

world in turn containing many Sims. However, our likelihood-ratio argument can apply to this case too. All we need assume is that our chances of occupying a given simulated world will decline as the overall number of worlds rises. Again, this assignment of likelihoods follows from an economy assumption about the simulation process, i.e. simulation costs are proportional to the number of worlds simulated as well as the number of levels in our hierarchy. Thus, besides the vertical ‘few ancestors’ UC, we can also run a horizontal ‘few contemporaries’ UC against the hypothesis that this world is but one of many simulated worlds. As with any numbers for our simulating ancestral levels, any numbers for our simulated contemporary worlds must be speculative. UC only applies where we are not invited to discount any of our existing empirical basis, and UC fails against standard birth-rank DA. However, our existing evidential basis could be exactly the same whether or not there are other Sim-levels or Sim-worlds connected to ours.²⁰

Doesn’t this argument oppose other ways of ramifying the structure of reality and so create its own ‘Presumptuous Philosopher’ counter-examples? Consider John Wheeler’s ‘oscillating cosmos’, wherein the universe undergoes a cycle of expansions and contractions. Each phase is sealed off by a terminal singularity which acts as Big Crunch at one phase’s end and Big Bang at the beginning of the new. If Nature’s constants get ‘scrambled’ at each singularity, the Anthropic Principle suggests we are overwhelmingly likely to find ourselves observing only cycles which contain conditions suitable for our evolution. However, recycling requires that each cycle possesses sufficient gravitational force to induce re-collapse. Any cycle with insufficient gravity to generate a terminal singularity must lack descendents, so any cycles before ours must have been closed. Can’t we adapt UC to this case? If we have no reason to think our cycle is unusual *qua* future terminus, and most levels have successors and ancestors, we probably shouldn’t expect our world to be open, i.e.

²⁰Leslie thinks DA’s force greatly reduces if the world is significantly indeterministic, since then birth-rank data might not furnish clues to our probable future. Even if indeterminism threatens standard DA, our anti-simulation DA might still succeed *ad hominem* against Bostrom. Unlike the diachronic DA, our anti-simulation argument is synchronic, and the simulation-process implies that the world would have to be profoundly indeterministic before the number of simulation-levels became indeterminate or noticeably acausal. (Assuming the number of levels isn’t deliberately rigged to be sensitive to micro-events, *à la* Schrödinger’s cat.)

without a terminal singularity. (Our cycle *is* probably unusual in permitting observers. However, *pace* SIA, we've no reason to think observer-bearing worlds are more, or less, likely to be open.) Alternatively, if we think this world is open, we might invoke a version of UC against the 'many ancestral worlds' hypothesis. Such 'Presumptuous Philosopher' results seem too powerful to be plausible. However, our anti-Bostrom UC poses no threat to Wheeler cycles. We have no evidence that our world is a terminal cycle and no reason to accept any economy assumptions about the overall number of cycles. Also, the evidential situations in the two cases are different. By hypothesis, no causal information passes across singularities, so we cannot observe other Wheeler cycles. However, Sim-worlds and Sim-levels must be causally linked. While we couldn't expect to observe other Wheeler cycles, we need extra (*ad hoc*) explanation for why we observe no traces of our simulators, e.g. our simulators retrospectively wipe any untoward disclosures from Sim-consciousness.

VI) Conclusions.

UC's failure reminds us that DA is a subjectivist Bayesian argument, which uses a reasonably salient reference class and generalises from well-supported empirical data. Monton's DA, in circumventing birth-rank data, removes the chief evidential asymmetry separating DA from UC. The Solarian Corollary derives our 'probable' location purely from two relative population-sizes regardless of any cost to our empirical beliefs. The Simulation Argument is *not* amenable to such a quick despatch as SC and may not even have sceptical consequences. (If we are Sims, there are still tables and chairs; they just might not be quite what we ordinarily think they are.) However, SC does point one moral about deriving probabilities from population-sizes: don't look at population-sizes in isolation but consider them in conjunction with other background information. In the Simulation case, we have other information besides hypothetical proportions of Sim and non-Sims – we know we're not simulators. Bostrom gives no reasons for thinking non-simulating locations should out-number simulating locations. Pending such reasons, UC fits the Simulation case: all else being equal, we shouldn't assume we live in a Sim-hierarchy. In this (unusual) case, we can reasonably favour 'few ancestors' and 'few contemporaries' hypotheses, because we have no empirical evidence that any ancestral levels or contemporary simulated worlds exist and, if such exist, our location is unlikely.

A common objection to Doomsday, Ussherian, Solarian and Simulation arguments might run thus: choice of reference class is not a purely logical matter but can be assessed (at least partly) by the success of its applications and its coherence with background metaphysical and epistemic assumptions. If a given reference class leads to deeply counter-intuitive conclusions, discounts existing empirical evidence or requires controversial metaphysical assumptions, we have good reason for at least scrutinising that reference class. Of the four arguments above, only DA uses a reasonably robust reference class that squares with our background evidence and general epistemic assumptions. DA at least projects from plausible assumptions to a conclusion that fits with our existing beliefs and inductive strategies.

Contra DA, Chambers (p. 447) protests that he can't be considered a random human because his birth-rank must exceed his father's, whereas a truly random draw would give the same probability to his preceding his father and to his father preceding him. Inevitably, Chambers' birth-rank will receive different probabilities if he considers himself *qua* random human or *qua* his father's son. Like any probabilistic argument, DA's conclusion can change with the reference class under which its explanandum is considered. Trying to establish which reference classes are truly salient is beyond this paper's scope but for many purposes the 'random human' reference class seems more salient than (e.g.) the 'random carbon-based object' reference class. However, DA assumes that the 'human' reference class is not only salient now but is salient throughout history. It isn't clear that this reference class is enduringly salient, nor that its synchronic identity- or boundary-conditions are unproblematic.

Seemingly, what makes a reference-class salient (or not) is its mapping of causal-nomological structures or significant patterns of co-occurring properties. Grouping red objects into one reference class makes sense for certain purposes because they all share some causal (e.g. optical) properties in common. Likewise, electrons make a truly robust reference-class because (we hypothesise) all electrons possess certain shared causal/nomological features, e.g. all electrons co-instantiate the same charge and the same (rest) mass. Considering the physiological and cognitive factors that shape human identity and self-awareness, we might well wonder if the set of all humans (past, present and future) forms a truly salient reference class in the way that the set of electrons might. Likewise, Sims and non-Sims need not occupy the same

reference class even if they are internally experientially indistinguishable. The situations of Sims in simulated worlds and non-Sims in unsimulated worlds will reflect different nomological structures and support different counterfactual inferences. Rationality does not demand that we treat the ‘human’ reference class as enduringly salient or that the same reference class covers Sims and non-Sims. Genuinely salient reference classes map robust causal and/or nomological properties but the above arguments offer conclusions no more salient or robust than their contrived reference classes. Unless or until a truly robust procedure for choosing reference-classes is forthcoming, we can remain unmoved by Doomsday arguments.

Philosophy

School of Philosophy, Psychology and Language Sciences

David Hume Tower

George Square

Edinburgh

EH8 9JX

A.Richmond@ed.ac.uk